

# HR Analysis & Performance Insights

## Project Overview

This project focuses on analyzing HR data to understand employee performance, attrition, and workforce distribution. The main goal is to identify why employees leave the company and which departments, job roles, or locations are most affected.

A Power BI dashboard was created to show key HR metrics such as total employees, attrition rate, employee exits by department, performance level, and job role. The dashboard helps HR teams quickly understand trends and make better decisions to improve employee retention and performance.

## Dataset Description

This dataset contains HR employee information such as department, job role, gender, performance score, engagement score, start date, exit date, and attrition status.

The data was cleaned and analyzed using Python, SQL, and Power BI to build an interactive HR dashboard.

## Exploratory Data Analysis using Python

Data preparation and cleaning were performed using Python:

### 1. Data Loading

Loaded the HR dataset into Python using pandas.

### 2. Data Understanding

Checked dataset structure, column names, and data types using df.info() and df.head().

```
[8]: df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3150 entries, 0 to 3149
Data columns (total 37 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Employee ID     3150 non-null    int64  
 1   EmployeeFirst 3150 non-null    object  
 2   ExitDate       3150 non-null    object  
 3   Title          3150 non-null    object  
 4   SupervisorID  3150 non-null    object  
 5   Email          3150 non-null    object  
 6   BusinessUnit   3150 non-null    object  
 7   EmployeeStatus 3150 non-null    object  
 8   EmployeeType   3150 non-null    object  
 9   EmployeeName   3150 non-null    object  
 10  EmployeeClassificationType 3150 non-null    object  
 11  TerminationType 3150 non-null    object  
 12  TerminationDescription 3150 non-null    object  
 13  DepartmentType 3150 non-null    object  
 14  Division        3150 non-null    object  
 15  DOB            3150 non-null    object  
 16  DOD            3150 non-null    object  
 17  Site           3150 non-null    object  
 18  JobFunctionDescription 3150 non-null    object  
 19  GenderCode      3150 non-null    object  
 20  LocationCode    3150 non-null    int64  
 21  MaritalDesc     3150 non-null    object  
 22  Performance Score 3150 non-null    object  
 23  Current Employee Rating 3150 non-null    int64  
 24  Survey ID       3150 non-null    int64  
 25  Survey Date     3150 non-null    object  
 26  Engagement Score 3150 non-null    int64  
 27  Satisfaction Score 3150 non-null    int64  
 28  Work-Life Balance Score 3150 non-null    int64  
 29  TotalScore      3150 non-null    int64 
```

### 3. Summary Statistics

Used df.describe() to understand distributions of numerical columns like age, scores, and ratings.

```
jupyter Python Last Checkpoint: 2 days ago
File Edit View Run Kernel Settings Help Trusted
+ × Code
In [9]: df = pd.read_csv('HR_comma_sep.csv')
        df['Training Cost'].hist()
        df.describe()
        df.info()
        df['Age'].hist()
        df['Score'].hist()
        df['Rating'].hist()
        df['Satisfaction Score'].hist()

Out[9]:
   EmployeeID StartDate ExitDate Title Supervisor ADEmail BusinessUnit EmployeeStatus EmployeeType PayZone ... SatisfactionScore
count      3150.000000      3150     1606      3150      3150      3150      3150      3150      3150      3150 ... 3150.000000
unique       NaN        1496      820       32      2952      2998       10         5         3         3 ...       NaN
top          NaN  18-Jul-19  21-Jul-23  ProductionTechnician I Paul Smith  uriah.bridges@bilearner.com      SVG      Active  Full-Time  Zone A ...       NaN
freq         NaN          8          10        1380         3          2        327      2583      1086      1108 ...       NaN
mean      2502.748254      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN ... 3.021270
std       866.241991      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN ... 1.407413
min      1001.000000      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN ... 1.000000
25%      1750.250000      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN ... 2.000000
50%      2504.500000      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN ... 3.000000
75%      3254.750000      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN ... 4.000000
max      4000.000000      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN ... 5.000000
11 rows × 37 columns
```

### 4. Missing Value Check

Identified missing values and handled them appropriately to maintain data quality.

### 5. Data Cleaning

Renamed columns to a standard format and fixed inconsistent values.

### 6. Feature Engineering

Created new columns such as age, age\_group, tenure\_group, attrition, and performance\_level.

### 7. Data Validation

Verified the correctness of newly created columns.

### 8. Data Export

Loaded the cleaned dataset into PostgreSQL for SQL analysis and Power BI reporting.

## Data Analysis using SQL

We performed structured analysis in PostgreSQL to answer key HR business questions:

1. **Total Employees Who Left** – Counted the total number of employees who have exited the company.

2. **Department-wise Attrition** – Identified departments with the highest attrition rate.
3. **Gender-based Attrition** – Compared whether men or women are leaving the company more frequently.
4. **Location-wise Attrition** – Found offices or locations with the highest employee exits.
5. **Average Tenure Before Exit** – Calculated the average time employees stayed before leaving the company.
6. **New Joinees in Last Year/Quarter/Month** – Counted how many employees joined in recent periods.
7. **Performance Level vs. Attrition** – Checked if low-performing employees leave more than high-performing ones.
8. **Current Employees by Department** – Counted the number of active employees in each department.
9. **Job Roles with High Attrition** – Identified job roles or titles that experience the highest turnover.
10. **Supervisors with High Attrition Teams** – Found supervisors whose teams have the highest attrition rate.

## HR Analytics Dashboard – Power BI

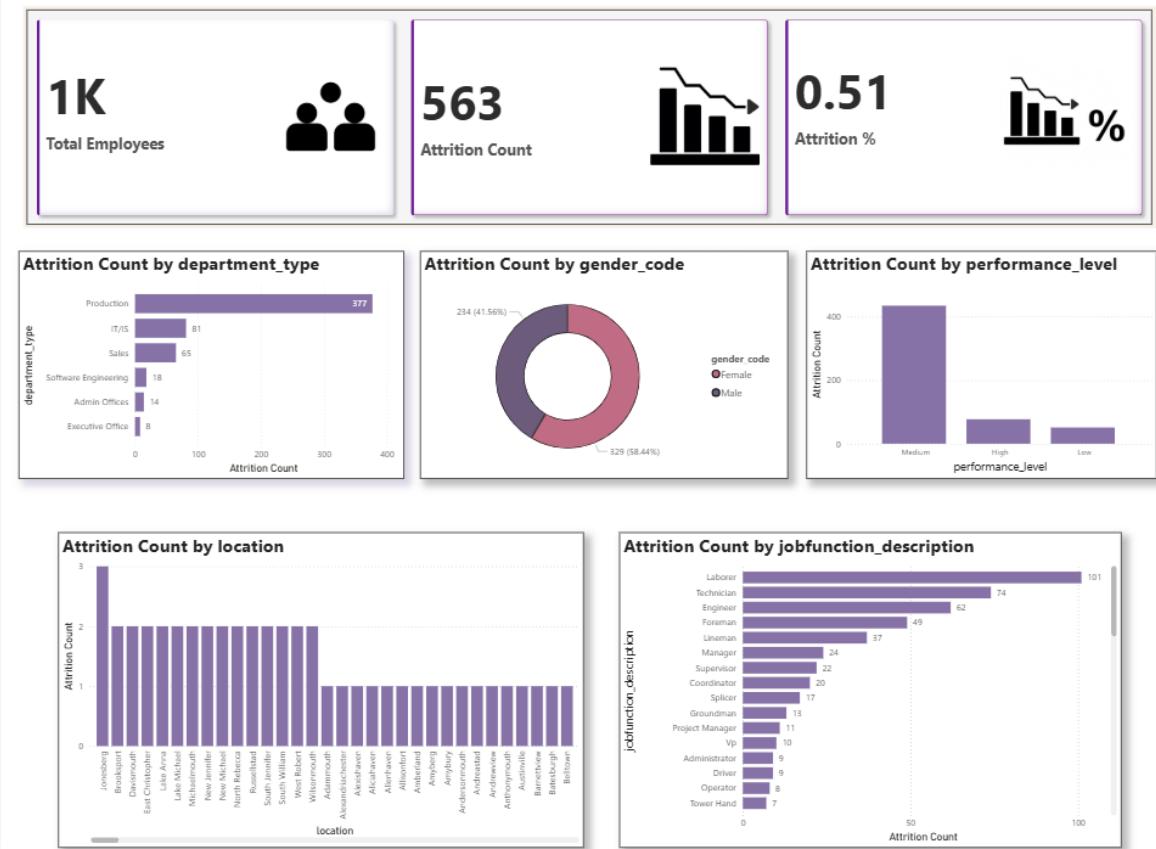
### Overview

Created an interactive HR dashboard in Power BI to help HR teams monitor workforce trends and make data-driven decisions.

### Key Highlights

- **Employee Overview:** Shows total employees, number of exits, attrition rate, and average tenure.
- **Attrition Insights:**
  - Departments with highest exits
  - Male vs Female attrition
  - Performance level impact on attrition
- **Workforce Details:**
  - High-risk job roles
  - Locations with more employee exits

# HR Analysis & Performance Dashboard



## Business Recommendations

- Improve retention strategies in high-attrition departments
- Focus on employee engagement for low-performing groups
- Strengthen retention programs for early-tenure employees
- Monitor location-wise attrition trends regularly

## Conclusion

The HR Analysis & Performance Dashboard provides a clear view of employee attrition, performance trends, and workforce distribution. Using Python, SQL, and Power BI, key HR metrics were analyzed and visualized to identify high-risk departments, job roles, and locations. This dashboard enables HR teams to make data-driven decisions for improving employee retention, performance management, and overall workforce planning.